Common dynamic estimation via structured low-rank approximation with multiple rank constraints

Antonio Fazzi * **Nicola Guglielmi** ** **Ivan Markovsky** *
Konstantin Usevich ***

* *Vrije Universiteit Brussel (VUB), Department ELEC, Pleinlaan 2, 1050 Brussels, Belgium (e-mail: antonio.fazzi@vub.be, ivan.markovsky@vub.be).*
** *Gran Sasso Science Institute (GSSI), Viale F. Crispi 7, 67100 L'Aquila, Italy (e-mail: nicola.guglielmi@gssi.it)*
*** *Université de Lorraine and CNRS, CRAN, F-54000 Nancy, France, (e-mail: konstantin.usevich@univ-lorraine.fr)*

Abstract: We consider the problem of detecting the common dynamic among several observed signals. It has been shown in (Markovsky et al., 2019) that the problem is equivalent to a generalization of the classical Hankel low-rank approximation to the case of multiple rank constraints. We propose an optimization method based on the integration of ordinary differential equations describing a descent dynamic for a suitable functional to be minimized. We show how the proposed algorithm improves the numerical solutions computed by existing subspace methods which solve the same problem.

Keywords: Common dynamics, Behavioral approach, Structured low-rank approximation, Data-driven estimation, Optimization problem

1. INTRODUCTION AND PROBLEM FORMULATION

The common dynamic estimation problem is defined as follows

Problem 1. Given a set of N scalar autonomous linear time-invariant systems $\mathcal{B}_1, \dots, \mathcal{B}_N$, find their common dynamics, defined as $\mathcal{B} := \mathcal{B}_1 \cap \dots \cap \mathcal{B}_N$.

The common dynamics estimation problem occurs in data-driven structured noise filtering (Markovsky et al., 2020), biomedical signal processing (Van Huffel, 1993; De Clercq et al., 2005), monitoring of material structures (Rippert, 2005), and audio modeling (Boyer and Abed-Meraim, 2004; Haneda et al., 1994). Subspace methods for common dynamics estimation are proposed in (Papy et al., 2006; Markovsky et al., 2019, 2020).

We work in the behavioral setting (Polderman and Willems, 1998; Willems, 2007; Markovsky et al., 2006), a mathematical theory where a dynamical system is defined as a set of signals instead of (parameter dependent) equations describing a model. The notation $y \in \mathcal{B}$ means that the signal y is a trajectory of the system $\mathcal{B}$. We consider the class of autonomous discrete-time linear time-invariant systems. A system $\mathcal{B}$ is linear if $\mathcal{B}$ is a subspace and it is time-invariant if $\sigma\mathcal{B} = \mathcal{B}$, where σ is the unit shift operator $(\sigma y)(t) = y(t+1)$ (acting on a set of signals σ shifts all signals in the set). The dimension of an autonomous linear time-invariant system is equal to its order. The class of autonomous linear time-invariant systems with order at most n is denoted by $\mathcal{L}_n$.

In this framework, the common dynamic of the systems $\mathcal{B}_1, \dots, \mathcal{B}_N$ is the system $\mathcal{B} = \mathcal{B}_1 \cap \dots \cap \mathcal{B}_N$. A data-driven common dynamic estimation problem is: given trajectories $y^1, \dots, y^N$ of $\mathcal{B}_1, \dots, \mathcal{B}_N$, respectively, find the common dynamic system $\mathcal{B}$. If the signals y^i are inexact (for example they are measured with additive noise or are not generated by linear time-invariant systems of bounded order), generically, the solution of the data-driven common dynamic problem is the trivial system $\mathcal{B} = \{0\}$. In this case, we consider the problem of approximate data-driven common dynamic, where a bound on the order of the common dynamic system is imposed.

In this paper, we consider an optimization problem that defines the maximum likelihood estimator for the common dynamic, under the assumption of zero mean, uncorrelated, Gaussian measurement noise. Given (noisy) trajectories $y^1, \dots, y^N$, orders $n_1, \dots, n_N$ and lower bound n for the order of the common dynamic system, the problem we solve is

$$\begin{aligned} &\text{minimize } \sqrt{\sum_{i=1}^{N} \|y^i - \hat{y}^i\|_2^2} \\ &\text{over } \hat{y}^1, \dots, \hat{y}^N, \hat{\mathcal{B}}, \hat{\mathcal{B}}_1, \dots, \hat{\mathcal{B}}_N \\ &\text{subject to } \hat{y}^i \in \hat{\mathcal{B}}_i \in \mathcal{L}_{n_i} \quad i = 1, \dots, N \\ &\quad \text{and } \hat{\mathcal{B}} = \hat{\mathcal{B}}_1 \cap \dots \cap \hat{\mathcal{B}}_N \in \mathcal{L}_n. \end{aligned} \tag{1}$$

As shown in (Markovsky et al., 2020), the maximum likelihood estimator problem (1) is equivalent to a generalized Hankel structured low-rank approximation problem with multiple rank constraints. Given a finite signal $y \in \mathbb{R}^T$, the Hankel matrix with n rows generated by y is

$$\mathcal{H}_n(y) := \begin{bmatrix} y(1) & y(2) & \cdots & y(T-n+1) \\ y(2) & y(3) & \cdots & y(T-n+2) \\ y(3) & y(4) & \cdots & y(T-n+3) \\ \vdots & \vdots & & \vdots \\ y(n) & y(n+1) & \cdots & y(T) \end{bmatrix}. \tag{2}$$

With this notation, (1) is equivalent to

$$\begin{aligned} &\text{minimize over } \hat{y}_1, \dots, \hat{y}_N \ \sqrt{\sum_{i=1}^{N} \|y_i - \hat{y}_i\|^2} \\ &\text{subject to} \\ &\text{rank } \mathcal{H}_{m+1}(\hat{y}_i) \le m \ \text{ for } \ i = 1, \dots, N \\ &\text{rank } (\mathcal{H}_{M+1}(\hat{y}_1), \cdots, \mathcal{H}_{M+1}(\hat{y}_N)) \le M, \end{aligned} \tag{3}$$

where $m = n_s + n_d, M = Nn_d + n_s$ and n_s, n_d are the number of common poles and distinct poles, respectively. As it happens for classical structured low-rank approximation problems, this is a nonconvex optimization problem for which there are no analytic solutions or standard solution methods. Observe that we have N time series and $N+1$ constraints, therefore standard structured low-rank approximation algorithms cannot be used for the considered problem. The algorithm in (Fazzi et al., 2021) is generalized in the following to solve the common dynamic estimation problem, i.e., to deal with block-Hankel structured matrices and with multiple rank constraints.

2. SOLUTION METHOD

We are going to adapt the algorithm in (Fazzi et al., 2021) to the considered problem since the presence of the last *coupling* constraint does not allow the use of standard Hankel low-rank approximation algorithms. The proposed algorithm seeks to minimize all the singular values of the involved structured matrices by moving them along a descent direction. This is done by combining the corresponding gradient systems associated with each singular value. We describe the details in the following.

For the sake of simplicity we consider two time series y_1, y_2 coming from two linear time invariant models of the same order $m = n_s + n_d$, but the ideas can be naturally generalized to any number of time series and to different orders. We collect the time series in a vector $y = (y1; y2)$. Assuming the vector y is noisy, the data of the problem are three full rank matrices $H_1(y_1) = \mathcal{H}_{m+1}(y_1), H_2(y_2) = \mathcal{H}_{m+1}(y_2), H_3(y_1, y_2) = (\mathcal{H}_{M+1}(y_1), \mathcal{H}_{M+1}(y_2))$. In order to have a solution to problem (3), we need the three matrices H_1, H_2, H_3 to be rank deficient.

Definition 1. We denote by $\sigma(H)$ the smallest singular value of the matrix H.

The idea is to perturb the parameter vector y until its components satisfy the constraints in (3). This is done by considering a perturbed vector of the form $\tilde{y} = y + \epsilon w$, where ϵ is a scalar denoting the total norm of the perturbation, while $w = (w1; w2)$ is a norm 1 vector (of the same dimension of y). In order to satisfy all the rank constraints we need that all the singular values $\sigma_1(\tilde{y}_1) = \sigma(H_1(\tilde{y}_1)), \sigma_2(\tilde{y}_2) = \sigma(H_2(\tilde{y}_2)), \sigma_3(\tilde{y}_1, \tilde{y}_2) = \sigma(H_3(\tilde{y}_1, \tilde{y}_2))$ vanish at the same time, or equivalently their sum $\sigma_1(\tilde{y}_1) + \sigma_2(\tilde{y}_2) + \sigma_3(\tilde{y}_1, \tilde{y}_2)$ vanishes. Observe that we have three matrices depending on two vectors, therefore a standard low rank approximation approach for the three different matrices could not work because we have fewer parameters than matrices.

The sought perturbation is computed using the following double iteration strategy:

- for a fixed value of ϵ (the norm of the perturbation vector) we seek to compute the vector $w = (w1; w2)$ whose components minimize the functional $\sigma_1(\tilde{y}_1) + \sigma_2(\tilde{y}_2) + \sigma_3(\tilde{y}_1, \tilde{y}_2)$;
- given the perturbation vector w, we increase the value of ϵ until we reach an admissible solution.

Remark 1. Since we are dealing with a numerical algorithm, it is enough to minimize the objective functional (the sum of the singular values) up to a small fixed threshold. Then we can use one of the subspace methods in (Markovsky et al., 2019) to make the Hankel matrices rank-deficient.

2.1 Optima perturbation vectors

In this section we consider ϵ as fixed and we seek to compute the (local) optimum perturbation vector w which minimizes the sum of the singular values $\sigma_1(\tilde{y}_1) + \sigma_2(\tilde{y}_2) + \sigma_3(\tilde{y}_1, \tilde{y}_2)$. This computation is done by looking for the steepest descent direction for the objective functional. The following standard result about eigenvalues perturbation for positive semidefinite matrices (Kato, 1995) is needed:

Lemma 1. Let $D(t)$ be a differentiable real symmetric matrix function for t in a neighborhood of 0, and let $\lambda(t)$ be an eigenvalue of $D(t)$ converging to a simple eigenvalue λ_0 of $D_0 = D(0)$ as $t \to 0$. Let x_0 be a normalized eigenvector of D_0 associated to λ_0. Then the function $\lambda(t)$ is differentiable near $t = 0$ with

$$\dot{\lambda} = x_0^\top \dot{D} x_0.$$

Lemma 1 can be adapted to derivatives of singular values (see (Fazzi et al., 2021, Section 3.2)). Since we will look for the derivative of a sum, we can then exploit the linearity of the derivative operator and split the sum into its three terms. The result of Lemma 1 is then applied to each of the singular values.

Consider the first singular value $\sigma_1(\tilde{y}_1)$, which depends only on the first time series. In the case of a scalar Hankel matrix we proceed as in (Fazzi et al., 2021). We have a perturbed matrix of the form $\tilde{H}_1 = \mathcal{H}_{m+1}(y_1 + \epsilon w_1) = \mathcal{H}_{m+1}(y_1) + \bar{\epsilon}\mathcal{H}_{m+1}(w_1)$. If we denote by u_1 and v_1 the left and right singular vectors associated with σ_1, remembering the link between eigenvalues and singular values, we have (omitting the dependence on both time and $\tilde{y}_1$)

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\sigma_1^2(\tilde{y}_1) &= v_1^\top \frac{\mathrm{d}}{\mathrm{d}t}\left(\tilde{H}_1^\top \tilde{H}_1\right) v_1 = 2\bar{\epsilon}\sigma u_1^\top \mathcal{H}_{m+1}(\dot{w}_1) v_1 \\ \dot{\sigma}_1 &= \bar{\epsilon} u_1^\top \mathcal{H}_{m+1}(\dot{w}_1) v_1. \end{aligned}$$

Therefore the optimal direction for $\sigma(\tilde{H}_1)$ is given by minimizing the function (up to constant terms)

$$\begin{aligned} &u_1^\top \dot{\mathcal{H}}_{m+1}(w_1) v_1 = \langle u_1 v_1^\top, \dot{\mathcal{H}}_{m+1}(w_1) \rangle = \\ &\langle P_{\mathcal{H}}(u_1 v_1^\top), \dot{\mathcal{H}}_{m+1}(w_1) \rangle, \end{aligned} \tag{4}$$

where the last equality follows from the fact that the matrix in the second term of the inner product is Hankel structured. The operator $P_{\mathcal{H}}(\cdot)$ denotes the orthogonal projection of the argument onto the Hankel structured matrices.

Remark 2. The projection of a rank one matrix onto the set of Hankel structured matrices is computed via the algorithm in (Korobeynikov, 2009, Section 5), which is faster than the classical antidiagonals averaging.

The point of minimum for the function in (4) is reached when the two arguments express the same direction but they have different signs. Therefore, the optimal direction for the first singular value is given by the stationary points of

$$\dot{E}_1 = -P_{\mathcal{H}}(u_1 v_1^\top). \tag{5}$$

Analogous statements hold true for the singular value $\sigma_2(\tilde{y}_2)$, which depends only on one paramater (the second time series). Calling u_2, v_2 the singular vectors of the matrix $\tilde{H}_2 = \mathcal{H}_{m+1}(y_2) + \bar{\epsilon}\mathcal{H}_{m+1}(w_2)$ associated with $\sigma(\tilde{H}_2)$, we have that the optimal direction is given by the stationary points of the equation

$$\dot{E}_2 = -P_{\mathcal{H}}(u_2 v_2^\top). \tag{6}$$

But the third matrix $\tilde{H}_3$, has a different structure, i.e., it is a block matrix whose blocks are Hankel structured. Hence some changes are needed. Let u_3, v_3 be the left and right singular vectors of $\tilde{H}_3$ associated with its smallest singular value. The difference with respect to the previous case is that we have to split the rank one matrix $u_3 v_3^\top$ into two different blocks,

$$u_3 v_3^\top = (u_3 v_4^\top, u_3 v_5^\top)$$

where the vector (v_4, v_5) is a partition of the vector v_3 according with the dimensions of the blocks of the matrix $(\mathcal{H}_{M+1}(\tilde{y}_1), \mathcal{H}_{M+1}(\tilde{y}_2))$. Since we want to decrease the smallest singular value, we still need to minimize a function of the form $u_3^\top(\mathcal{H}_{M+1}(\dot{w}_1), \mathcal{H}_{M+1}(w_2))v_3$. The expression for the derivative of the smallest singular value, as in the other cases, is given by (up to constant terms)

$$\dot{\sigma}_3 = u_3^\top \frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{H}_{M+1}(w_1), \mathcal{H}_{M+1}(w_2))v_3.$$

Therefore such a derivative reaches the point of minimum in correspondence of the stationary points of the equation

$$\dot{E}_3 = -(P_{\mathcal{H}}(u_3 v_4^\top), P_{\mathcal{H}}(u_3 v_5^\top)), \tag{7}$$

and the corresponding perturbations on the time series are given by the two different blocks.

Given the two time series $\tilde{y}_1, \tilde{y}_2$, the vector $\tilde{y} = (\tilde{y}_1; \tilde{y}_2) = y + \epsilon w$, and the three pairs of singular vectors u_1, v_1, u_2, v_2 and u_3, v_3 associated with the smallest singular values of the matrices $\tilde{H}_1, \tilde{H}_2, \tilde{H}_3$, respectively, we need to solve the following optimization problem

$$\begin{aligned} g_* = \arg \min_{\substack{\dot{w}=(\dot{w}_1,\dot{w}_2)\\ \|\dot{w}\|=1}} & u_1^\top \mathcal{H}_{m+1}(\dot{w}_1)v_1 + u_2^\top \mathcal{H}_{m+1}(\dot{w}_2)v_2 + \\ & + u_3(\mathcal{H}_{M+1}(\dot{w}_1), \mathcal{H}_{M+1}(\dot{w}_2))v_3 \\ \text{subject to } & \dot{w}^\top \cdot w = 0. \end{aligned} \tag{8}$$

The stationary points of the three equations (5), (6) and (7) give three matrices $E_1, E_2, E_3 = (E_3^1, E_3^2)$ which contain the information on the directions for the perturbations on the time series $\tilde{y}_1, \tilde{y}_2$. The unconstrained solution of problem (8) is given by

$$\begin{aligned} \dot{w}_1 &= \mathbf{vec}(E_1) + \mathbf{vec}(E_3^1) \\ \dot{w}_2 &= \mathbf{vec}(E_2) + \mathbf{vec}(E_3^2) \end{aligned} \tag{9}$$

where the operator $\mathbf{vec}(\cdot)$ extract the first row and the last column of the argument (Hankel) matrix.

The time series $\dot{w}_1, \dot{w}_2$, however, represent the free gradient for the objective functional, so they do not satisfy the constraints in (8) and consequently they do not give an admissible solution for the problem. We need to satisfy the constraints

$$\|(\dot{w}_1, \dot{w}_2)\|_2 = 1; \quad (\dot{w}_1, \dot{w}_2)\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = 0.$$

The orthogonality condition is satisfied by building the following vector starting from the one in (9) $\dot{w} = (\dot{w}_1, \dot{w}_2)$:

$$\dot{w} = -\dot{w} + (\dot{w}^\top \cdot w)\, w.$$

This vector needs then to be divided by its norm in order to satisfy the constraint on the norm.

By summarizing all the previous results, we have that the optimal perturbation w is given by integrating the following system (possibly normalized):

$$\dot{w} = -F + \langle F, w \rangle w \tag{10}$$

where F is the vector whose components are the ones in the right hand side of (9).

2.2 Estimation of the time series

In the previous section we understood how to compute optimal perturbation vectors of given norm. The next step is to move the value ϵ of this norm until the objective functional reaches the point of (local) minimum.

First of all we need to consider a *free* dynamic for the objective functional, i.e., we want to minimize the objective functional without adding any constraint on the computed solution. Given the time series $\tilde{y} = y + \epsilon w$ and the three pairs of singular vectors $u_1, v_1, u_2, v_2, u_3, v_3$ associated with the smallest singular values of the three matrices $\tilde{H}_1, \tilde{H}_2, \tilde{H}_3$, compute

$$\begin{aligned} g_* = \arg \min_{\dot{w}_1, \dot{w}_2} & u_1^\top \mathcal{H}_{m+1}(\dot{w}_1)v_1 + u_2^\top \mathcal{H}_{m+1}(\dot{w}_2)v_2 + \\ & + u_3(\mathcal{H}_{M+1}(\dot{w}_1), \mathcal{H}_{M+1}(\dot{w}_2))v_3. \end{aligned} \tag{11}$$

The corresponding gradient system, following the same notation as in (10), is

$$\dot{w} = -F. \tag{12}$$

The difference with respect to the constrained case is that, during the integration of the system (12), the norm of the perturbation vector w increases while the objective functional decreases (for a fixed value of ϵ). This feature is exploited in the computation of an admissible solution by applying iteratively the following algorithm:

(1) Given the perturbed time series $\tilde{y} = y + \epsilon_k w$ at the k-th iteration and the value ϵ_{k+1}, integrate the equation (12) starting from the perturbation $e = \frac{\epsilon_k}{\epsilon k+1} w$, and stop when $\|e\| = 1$ (or equivalently $\epsilon_k \|w\|$ reaches the value ϵ_{k+1});
(2) solve the constrained problem (integration of (10)) with $\epsilon = \epsilon_{k+1}$ starting from the perturbation computed at the previous point.

Algorithm 1. Scheme of the main algorithm
Data: $y = (y_1; y_2)$ and tol (stopping tolerance)
Result: v (functional), $\sigma_1, \sigma_2, \sigma_3$, ϵ, $\tilde{y} = (\tilde{y}_1, \tilde{y}_2)$.
begin
  $H_1 = \mathcal{H}_{m+1}(y_1)$,
  $H_2 = \mathcal{H}_{m+1}(y_2)$,
  $H_3 = (\mathcal{H}_{M+1}(y_1), \mathcal{H}_{M+1}(y_2))$
  $\sigma_1 = \sigma(H_1), \sigma_2 = \sigma(H_2), \sigma_3 = \sigma(H_3), v = \sigma_1 + \sigma_2 + \sigma_3$
  epsilon=[], $\epsilon_0 = 10^{-2}$
  Integrate the ODE (10) with data $y = (y_1, y_2)$ and constrain the norm of the perturbation to ϵ_0
  Update $\sigma_1, \sigma_2, \sigma_3, v$ and store the perturbation w
  while $v > tol$ **do**
    epsilon=[epsilon, ϵ_0]
    $\epsilon_0 = \epsilon_0 + 10^{-2}$
    Integrate the equation (12) with inputs $\tilde{y} = y+$ epsval(end)$/\epsilon_0 * w$ and bound the perturbation to ϵ_0
    Update $\sigma_1, \sigma_2, \sigma_3, v$ and store the perturbation w
    Update $\epsilon_0 = \epsilon_0 * \|w\|_2$
    Integrate the equation (10) with inputs $\tilde{y} = y + \epsilon_0 w$ and constrain the norm of the perturbation to ϵ_0
    Update $\sigma_1, \sigma_2, \sigma_3, v$ and store the perturbation w
    if $|\epsilon_0$ - *epsilon(end)*$| < tol$ **then**
      STOP: found optimal ϵ
    end
  end
end

The previous algorithm allows at the first step to move from the set of perturbation vectors of norm ϵ_k to the set of perturbation vectors of norm ϵ_{k+1}. Then, in step 2, we compute an optimum (local) minimum for the functional by constraining the value of the norm of the perturbation. We iterate the algorithm until the functional reaches a fixed (small) tolerance. Alternatively we stop when we are not able anymore to improve the value of the objective functional by increasing the norm of the perturbation.

The outputs of the algorithm are a scalar ϵ which is the norm of the perturbation vector on the starting time series $y = (y_1, y_2)$ and a vector $w = (w_1, w_2)$ (of norm 1) containing the perturbation on y. The pseudocode for the algorithm is given in Algorithm 1.

3. NUMERICAL RESULTS

Some numerical results are given in order to understand the behavior of the proposed algorithm. We consider two discrete signals $y^t = (y_1^t, y_2^t)$ made of 50 points each (the superscript t denotes the true signals). They are two multisine waves coming from two systems of order 6 having 2 common exponents. The data for our experiments are then generated by adding a random perturbation to each element of the vectors, and in particular,

$$y = y^t * \sigma * r/\|r\|_2 * \|y^t\|_2,$$

where $\sigma = 0.25$ is a constant representing the level of noise while r is a random vector whose components come from a standard normal distribution.

By running the algorithm, we observe the objective functional (the sum of the three singular values) decreases from 0.645 to $1.78 \cdot 10^{-3}$, and in particular we have $\sigma_1 : 0.196 \to 9.3 \cdot 10^{-4}$, $\sigma_2 : 0.207 \to 1.2 \cdot 10^{-7}$, $\sigma_3 : 0.242 \to 7.3 \cdot 10^{-4}$ (Figure 1). We observe that the decrease is faster during the first iterations while it slows down as the iteration number increases. This is probably due to the choice of the explicit Euler scheme for the integration of the equation.

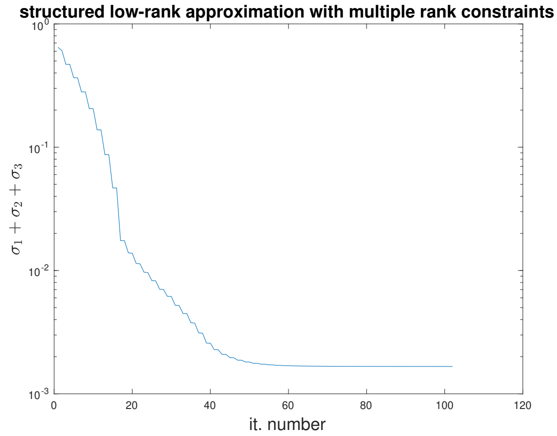

Fig. 1. Numerical solution of a Hankel low-rank approximation problem with multiple rank constraints: decreasing of the sum of the singular values.

The value of ϵ in output for the time series computed by the experiment in Figure 1 is 0.2227, while the true value (the difference between the true signals and the noisy signals) is 0.2928. However the considered tolerance for the singular values is 10^{-6}, so only one of the three matrices (the second) can be actually considered rank deficient: both the first and the third singular values are bigger than the fixed tolerance 10^{-6}. All the singular values decreased but the algorithm stopped because two successive values of ϵ are almost equal. In order to decrease all the singular values below the threshold 10^{-6} we can go on as follows: we fix the second time series (since the second singular value reached the fixed tolerance) and we go on perturbing the first time series only until all the constraints are satisfied (the numerical strategy is the same but the optimization is run on one parameter only). In this way the sum of the three singular values becomes smaller than 10^{-6}. The final distance between the computed and the starting time series is 0.4311 (so we get an upper bound with respect to the true value 0.2928), while the relative error is 0.016.

We compare finally this relative error with the one computed using the subspace method presented in (Markovsky et al., 2019) labeled as *papy-ls*, which is 0.3281. Therefore we got an improvement of one order of magnitude.

Remark 3. We remark that the algorithm computes an *approximate* common dynamic among the given (noisy) signals, so we expect to find a solution even if the intersection of the original noiseless signals is empty. At the moment, several properties of the algorithm (e.g., speed of convergence and estimates on the error on the computed solution) are only numerical evidence and still need to be investigated.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the Fond for Scientific Research Vlaanderen (FWO)

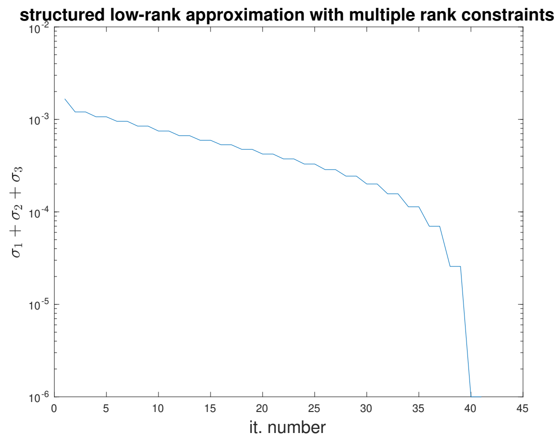

Fig. 2. Decreasing of the sum of the singular values in the numerical solution of a Hankel low-rank approximation problem with multiple rank constraints: optimization on one parameter vector only.

projects G028015N and G090117N and the FNRS-FWO under Excellence of Science (EOS) Project no 30468160 "Structured low-rank matrix / tensor approximation: numerical optimization-based algorithms and applications".

REFERENCES

Boyer, R. and Abed-Meraim, K. (2004). Audio modeling based on delayed sinusoids. *IEEE Transactions on Speech and Audio Processing*, 12(2), 110–120.

De Clercq, W., Vanrumste, B., Papy, J., Van Paesschen, W., and Van Huffel, S. (2005). Modeling common dynamics in multichannel signals with applications to artifact and background removal in EEG recordings. *IEEE Transactions on Biomedical Engineering*, 52(12), 2006–2015.

Fazzi, A., Guglielmi, N., and Markovsky, I. (2021). A gradient system approach for Hankel structured low-rank approximation. *Linear Algebra Appl.* doi: 10.1016/j.laa.2020.11.016.

Haneda, Y., Makino, S., and Kaneda, Y. (1994). Common acoustical pole and zero modeling of room transfer functions. *IEEE Transactions on Speech and Audio Processing*, 2(2), 320–328.

Kato, T. (1995). *Perturbation theory for linear operators.* Springer-Verlag.

Korobeynikov, A. (2009). Computation- and space-efficient implementation of ssa. *Stat. Interface*, (3), 357–368.

Markovsky, I., Liu, T., and Takeda, A. (2019). Subspace methods for multi-channel sum-of-exponentials common dynamics estimation. In *Proc. of the IEEE Conf. on Decision and Control*, 2672–2675. doi:978-1-7281-1397-5/19.

Markovsky, I., Liu, T., and Takeda, A. (2020). Data-driven structured noise filtering via common dynamics estimation. *IEEE Trans. Signal Process.*, 68, 3064–3073.

Markovsky, I., Willems, J.C., Van Huffel, S., and De Moor, B. (2006). *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach.* SIAM.

Papy, J.M., De Lathauwer, L., and Van Huffel, S. (2006). Common pole estimation in multi-channel exponential data modeling. *Signal Processing*, 86(4), 846–858.

Polderman, J. and Willems, J.C. (1998). *Introduction to Mathematical Systems Theory.* Springer-Verlag, New York.

Rippert, L. (2005). *Optical fibers for damage monitoring in carbon fiber reinforced plastic composite materials.* Ph.D. thesis, Katholieke Universiteit Leuven.

Van Huffel, S. (1993). Enhanced resolution based on minimum variance estimation and exponential data modeling. *Signal Processing*, 33(3), 333–355.

Willems, J.C. (2007). The behavioral approach to open and interconnected systems: Modeling by tearing, zooming, and linking. *Control Systems Magazine*, 27, 46–99.